CLAIMS

What is claimed is:

1. A method, comprising:

electronically capturing visual features associated with a speaker speaking;

electronically capturing audio;

matching selective portions of the audio with the visual features; and

identifying the remaining portions of the audio as potential noise not associated with the speaker speaking.

2. The method of claim 1 further comprising:

electronically capturing additional visual features associated with a different speaker speaking; and

matching some of the remaining portions of the audio from the potential noise with the additional speaker speaking.

3. The method of claim 1 further comprising generating parameters associated with the matching and the identifying and providing the parameters to a Bayesian Network which models the speaker speaking.

4. The method of claim 1 wherein electronically capturing the visual features further includes processing a neural network against electronic video associated with the speaker speaking, wherein the neural network is trained to detect and monitor a face of the speaker.

5. The method of claim 4 further comprising filtering the detected face of the speaker to detect movement or lack of movement in a mouth of the speaker.

6. The method of claim 1 wherein matching further includes comparing

portions of the captured visual features against portions of the captured audio during a same time slice.

7. The method of claim 1 further comprising suspending the capturing of audio during periods where select ones of the captured visual features indicate that the speaker is not speaking.

8. A method, comprising:

monitoring an electronic video of a first speaker and a second speaker;

concurrently capturing audio associated with the first and second speaker speaking;

analyzing the video to detect when the first and second speakers are moving their respective mouths; and

matching portions of the captured audio to the first speaker and other portions to the second speaker based on the analysis.

9. The method of claim 8 further comprising modeling the analysis for subsequent interactions with the first and second speakers.

10. The method of claim 8 wherein analyzing further includes processing a neural network for detecting faces of the first and second speakers and processing vector classifying algorithms to detect when the first and second speakers' respective mouths are moving or not moving.

11. The method of claim 8 further comprising separating the electronic video from the concurrently captured audio in preparation for analyzing.

12. The method of claim 8 further comprising suspending the capturing of audio when the analysis does not detect the mouths moving for the first and second speakers.

13. The method of claim 8 further comprising identifying selective portions of the captured audio as noise if the selective portions have not been matched to the first speaker or the second speaker.

14. The method of claim 8 wherein matching further includes identifying time dependencies associated with when selective portions of the electronic video were monitored and when selective portions of the audio were captured.

15. A system, comprising:

a camera;

a microphone; and

a processing device, wherein the camera captures video of a speaker and communicates the video to the processing device, the microphone captures audio associated with the speaker and an environment of the speaker and communicates the audio to the processing device, the processing device includes instructions that identifies visual features of the video where the speaker is speaking and uses time dependencies to match portions of the audio to those visual features.

16. The system of claim 15 wherein the captured video also includes images of a second speaker and the audio includes sounds associated with the second speaker, and wherein the instructions matches some portions of the audio to the second speaker when some of the visual features indicate the second speaker is speaking.

17. The system of claim 15 wherein the instructions interact with a neural network to detect a face of the speaker from the captured video.

18. The system of claim 17 wherein the instructions interact with a pixel

vector algorithm to detect when a mouth associated with the face moves or does not move within the captured video.

19. The system of claim 18 wherein the instructions generate parameter data that configures a Bayesian network which models subsequent interactions with the speaker to determine when the speaker is speaking and to determine appropriate audio to associate with the speaker speaking in the subsequent interactions.

20. A machine accessible medium having associated instructions, which when accessed, results in a machine performing:

separating audio and video associated with a speaker speaking;

identifying visual features from the video that indicate a mouth of the speaker is moving or not moving; and

associating portions of the audio with selective ones of the visual features that indicate the mouth is moving.

21. The medium of claim 20 further including instructions for associating other portions of the audio with different ones of the visual features that indicate the mouth is not moving.

22. The medium of claim 20 further including instructions for:

identifying second visual features from the video that indicate a different mouth of another speaker is moving or not moving; and

associating different portions of the audio with selective ones of the second visual features that indicate the different mouth is moving.

23. The medium of claim 20 wherein the instructions for identifying further include instructions for:

processing a neural network to detect a face of the speaker; and

processing a vector matching algorithm to detect movements of the

mouth of the speaker within the detected face.

24. The medium of claim 20 wherein the instructions for associating further include instructions for matching same time slices associated with a time that the portions of the audio were captured and the same time during which the selective ones of the visual features were captured within the video.

25. An apparatus, residing in a computer-accessible medium, comprising:

face detection logic;

mouth detection logic; and

audio-video matching logic, wherein the face detection logic detects a face of a speaker within a video, the mouth detection logic detects and monitors movement and non-movement of a mouth included within the face of the video, and the audio-video matching logic matches portions of captured audio with any movements identified by the mouth detection logic.

26. The apparatus of claim 25 wherein the apparatus is used to configure a Bayesian network which models the speaker speaking.

27. The apparatus of claim 25 wherein the face detection logic comprises a neural network.

28. The apparatus of claim 25 wherein the apparatus resides on a processing device and the processing device is interfaced to a camera and a microphone.